On the Relationship Between Geospatial Liquefaction-Model Performance and Quality of Geospatial Data: A Case Study of the 2010-2016 Canterbury Earthquakes

Alexander J. Baird[1], Mertcan Geyin[1], and Brett W. Maurer[1]

[1] Dept. of Civil and Environmental Engineering, University of Washington, USA

UNIVERSITY of WASHINGTON

ABSTRACT

Geospatial liquefaction models aim to predict liquefaction using data that is free and readily-available. This data includes (i) common ground-motion intensity measures; and (ii) geospatial parameters (e.g., among many, distance to rivers, distance to coast, and V_{s30} estimated from topography) which are used to infer characteristics of the subsurface without in-situ testing. Since their recent inception, such models have been used to predict geohazard impacts throughout New Zealand (e.g., in conjunction with regional ground-motion simulations). While past studies have demonstrated that geospatial liquefaction-models show great promise, the resolution and accuracy of the geospatial data underlying these models is notably poor. As an example, mapped rivers and coastlines often plot hundreds of meters from their actual locations. This stems from the fact that geospatial models aim to rapidly predict liquefaction anywhere in the world and thus utilize the lowest common denominator of available geospatial data, even though higher quality data is often available (e.g., in New Zealand). Accordingly, this study investigates whether the performance of geospatial models can be improved using higher-quality input data. This analysis is performed using (i) 15,101 liquefaction case studies compiled from the 2010-2016 Canterbury Earthquakes; and (ii) geospatial data readily available in New Zealand. In particular, we utilize alternative, higher-quality data to estimate: locations of rivers and streams; location of coastline; depth to ground water; V_{s30}; and PGV. Most notably, a region-specific V_{s30} model improves performance (Figs. 3-4), while other data variants generally have little-to-no effect, even when the "standard" and "high-quality" values differ significantly (Fig. 2). This finding is consistent with the greater sensitivity of geospatial models to V_{s30}, relative to any other input (Fig. 5), and has implications for modeling in locales worldwide where high quality geospatial data is available.

DATA & METHODOLOGY

Model performance is assessed for 15,101 case-studies compiled from three earthquakes (Table 1). The study locations are mapped in Figure 1. The geotechnical aspects of this dataset are discussed in detail in [1].

Table 1. Case Studies Analyzed

Mo/Year	Earthquake	# Cases
Sept 2010	Darfield	5,418
Feb 2011	Christchurch	4,847
Feb 2016	Valentines Day	4,836
		15,101

Figure 1. Case-study locations

Geospatial Models

The models of Zhu et al.[5,6], computed from geospatial data as:

$$P(X) = (1 + e^{-X})^{-1}$$

1. **Global Geospatial Model 1 – Coastal Zones** (Zhu et al., 2017)[6]:
$$X = 12.435 + 0.301\cdot\ln(PGV) - 2.615\cdot\ln(V_{s30}) + 5.556 \times 10^{-4}\cdot precip - 0.0287\cdot(d_c)^{0.5} + 0.0666\cdot d_r - 0.0369\cdot d_r\cdot(d_c)^{0.5}$$

2. **Global Geospatial Model 2 – Noncoastal Zones** (Zhu et al., 2017)[6]:
$$X = 8.801 + 0.334\cdot\ln(PGV) - 1.918\cdot\ln(V_{s30}) + 5.408 \times 10^{-4}\cdot precip - 0.2054\cdot d_w - 0.0333\cdot wtd$$

3. **Global Geospatial Model** (Zhu et al., 2015)[5]:
$$X = 24.10 + 2.067\cdot\ln(PGA_M) + 0.355\cdot CTI - 0.4784\cdot\ln(V_{s30})$$

4. **Christchurch Geospatial Model 1** (Zhu et al., 2015)[5]:
$$X = 2.053 + 1.267\cdot\ln(PGA_M) - 0.239\cdot d_{r3} - 9.191\cdot ND$$

5. **Christchurch Geospatial Model 2** (Zhu et al., 2015)[5]:
$$X = 0.316 + 1.225\cdot\ln(PGA_M) + 0.145\cdot CTI - 9.708\cdot ND$$

6. **Christchurch Geospatial Model 3** (Zhu et al., 2015)[5]:
$$X = 25.45 + 2.476\cdot\ln(PGA_M) - 0.323\cdot d_{r3} - 4.241\cdot\ln(V_{s30})$$

Where: $P(X)$ = probability of surface manifestation; CTI = compound topographic index; PGA_M = magnitude-weighted peak ground acceleration; PGV = peak ground velocity (cm/s); V_{s30} = shear-wave velocity of the upper 30-m (m/s); d_r = distance to river (km); d_{r3} = distance to stream of order three or greater (km); d_c = distance to coast (km); $precip$ = mean annual precipitation (mm); ND = d_c divided by the distance from the coast to the edge of the sedimentary basin; wtd = water table depth (m).

"Standard" models #1-6 were computed using the data sources of Zhu et al. (2015; 2017). **"High Quality" models #1-6** were computed using alternative data sources/models, including: (a) *PGVs* from "Seisfinder" physics-based simulations[2]; (b) a New Zealand specific V_{s30} model[3]; (c) d_c measured from a coastline polygon[4]; (d) Digital Elevation Model (DEM) derived parameters (e.g., d_{r3}) using higher resolution DEMs (SRTM 1 arc-second and 8-meter DEMs)[4]. See Figure 2 for a comparison of inputs.

DATA COMPARISON

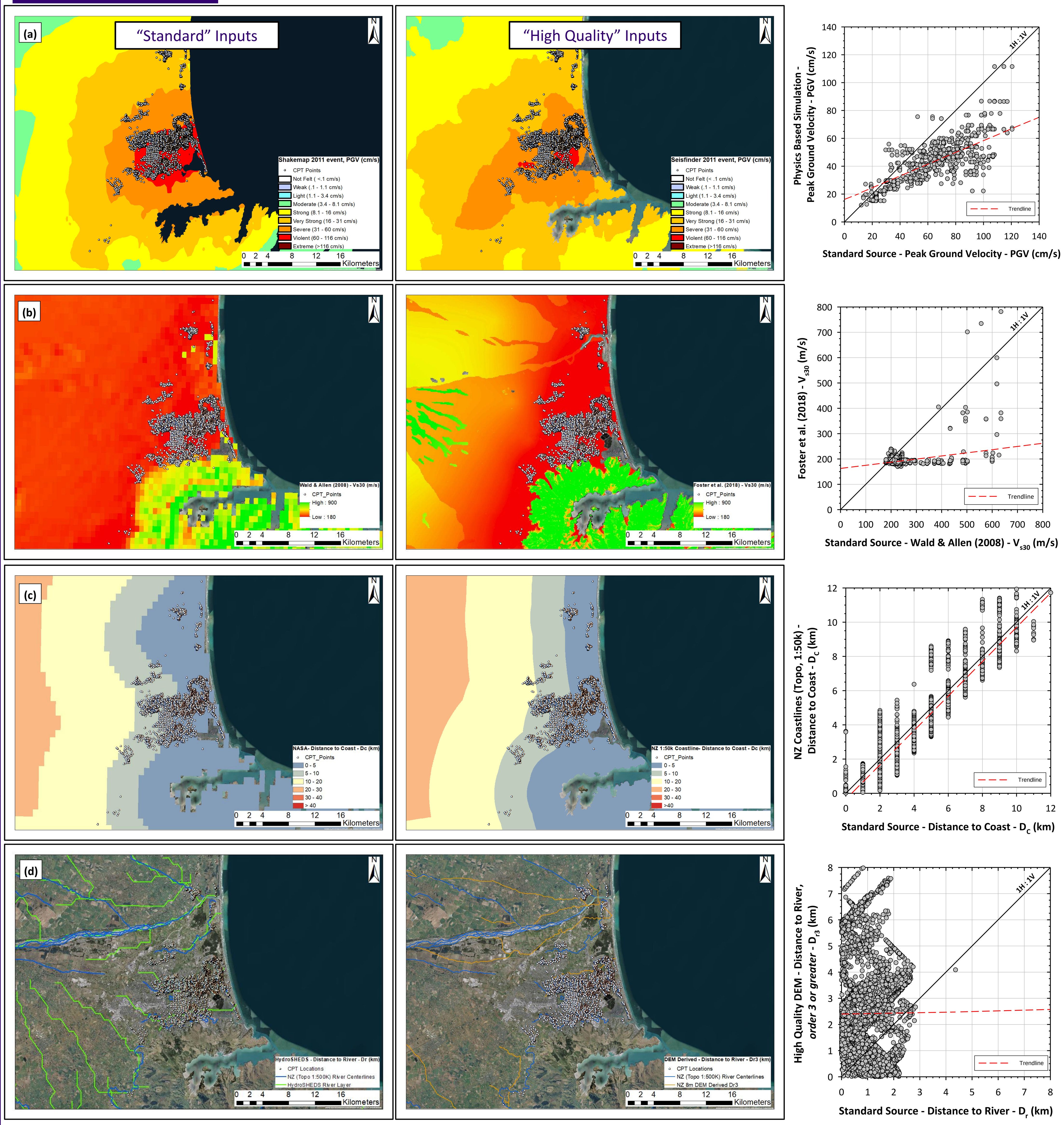

Figure 2. Comparison of standard geospatial model inputs (left) and higher quality alternatives (right) : (a) PGV (m/s); (b) V_{s30} (cm/s); (c) D_c (km); and (d) D_r (km).

RESULTS

Model efficacy is assesed using receiver-operating-characteristic (ROC) analyses, which plot the rates of true- and false-positive predictions as a function of index-test results. The area under the ROC curve (AUC) is used to characterize model efficiency and is the probability that "manifestation" cases have higher index-values than "no manifestation" cases. A larger AUC thus indicates better model performance.

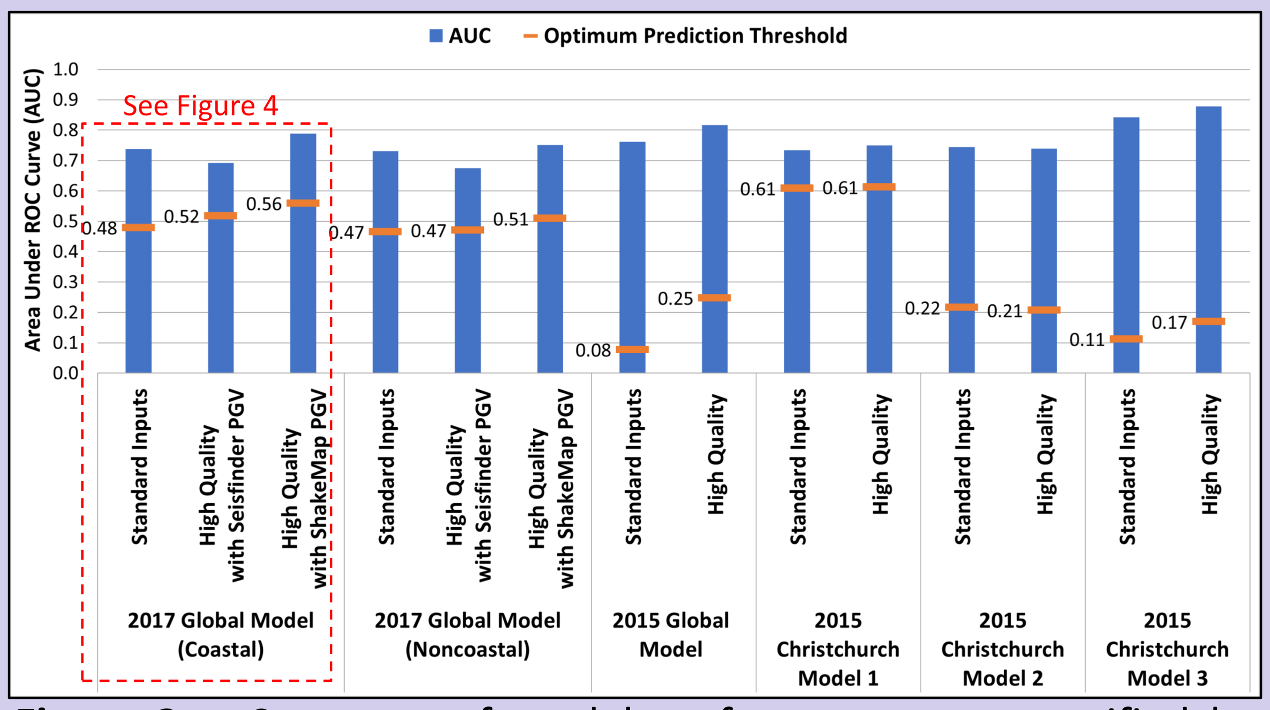

Figure 3. Summary of model performance, as quantified by AUC; the optimal threshold value is also shown for each model.

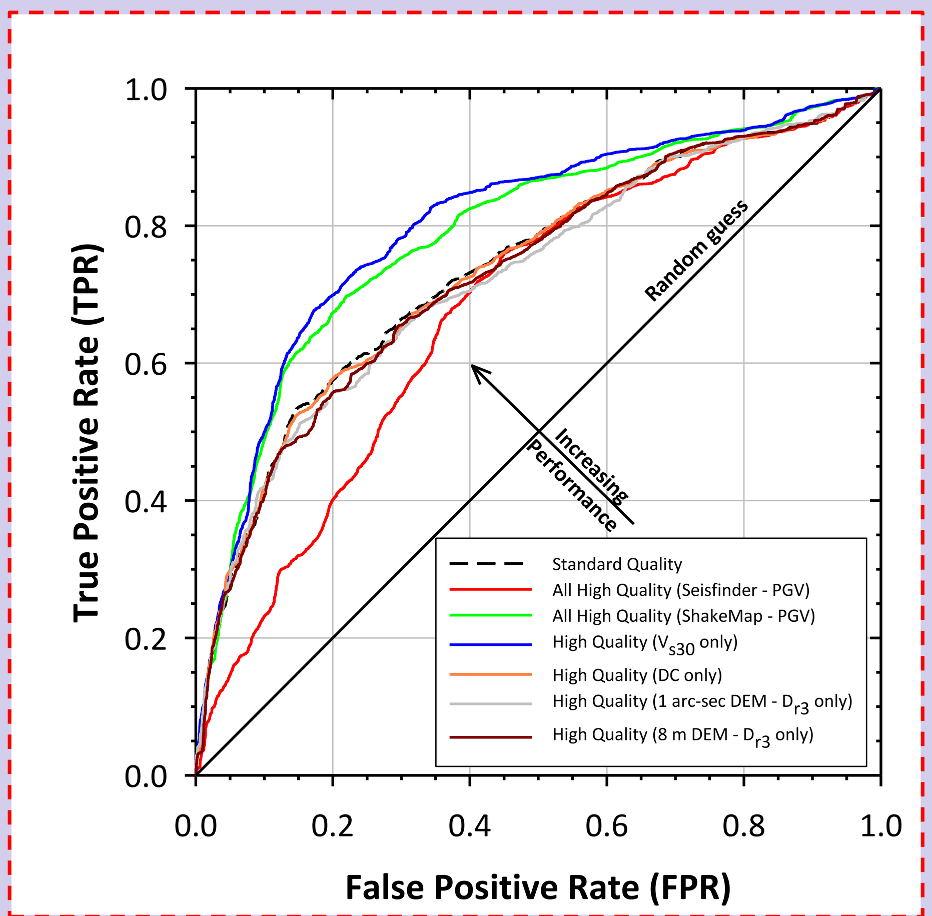

Figure 4. ROC curves showing influence of using "High Quality" data variants, both collectively and separately, in this case for the 2017 Global Model (Coastal).

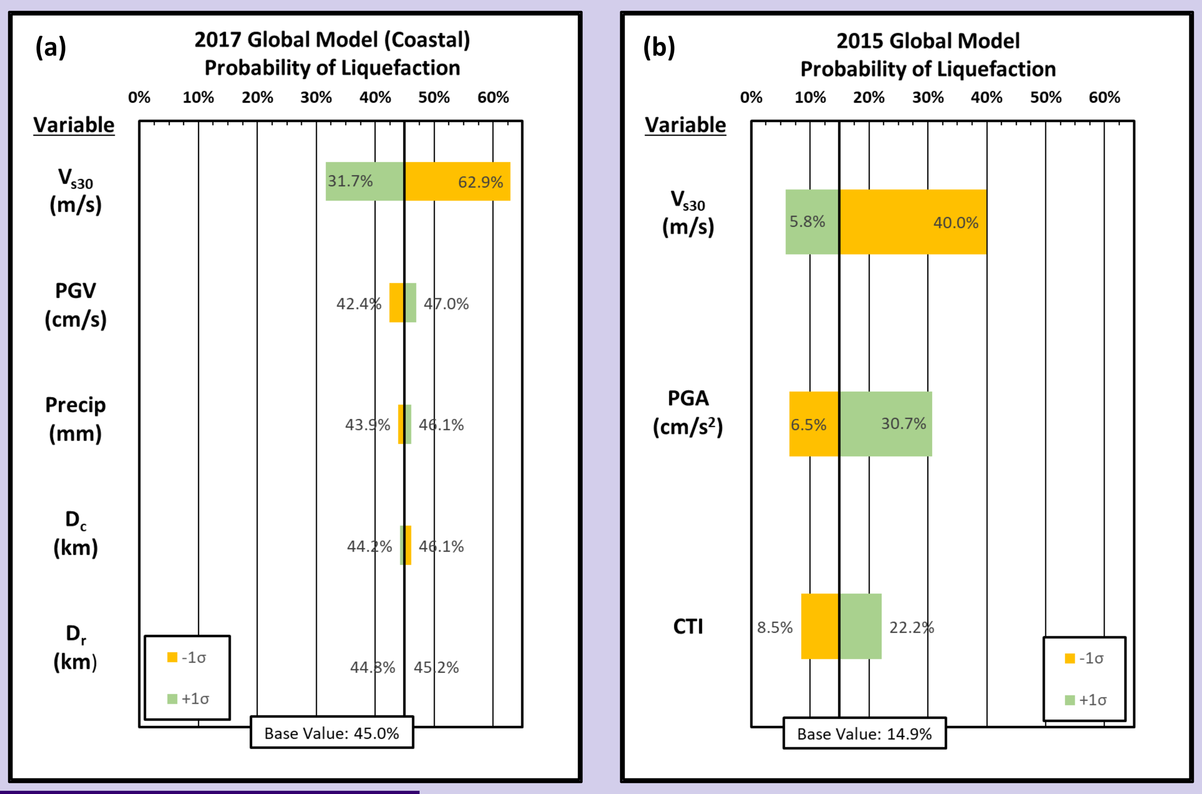

Figure 5. "Tornado" diagrams (Feb. 2011 event) showing sensitivity of predicted liquefaction probabilities to input-parameter variation (±1σ of population statistics): (a) 2017 Global Model (Coastal); and (2) 2015 Global Model.

KEY FINDING

A region-specific, high-resolution V_{s30} model improved performance (Figs. 3-4), while other data variants had little-to-no effect, even when the "standard" and "high-quality" values differed significantly (Fig. 2). This finding is consistent with the greater sensitivity of geospatial models to V_{s30}, relative to any other input (Fig. 5), and is due largely to improved predictions near the Port Hills. This suggests that estimating V_{s30} with greater accuracy and/or spatial resolution is particularly important for the geospatial liquefaction models assessed herein.

REFERENCES

[1] Maurer, B.W., Green, R.A., Cubrinovski, M., and Bradley, B. (2015). "Fines-content effects on liquefaction hazard evaluation for infrastructure in Christchurch, New Zealand." *Soil Dynamics and Earthquake Engineering* 76: 58-68.

[2] Savarimuthu S, Lagrava D, Bradley BA, Huang J, Motha J, Polak V, Bae S. (2017). "SeisFinder: A web application for extraction of data from computationally-intensive earthquake resilience calculations.: *QuakeCoRE Annual Meeting*. 3-6 September 2017. Taupo, New Zealand.

[3] Foster, K., Bradley, B., McGann, C., Wotherspoon, L. (2018). "A Vs30 Map for New Zealand based on Geologic and Terrain Proxy Variables, Updated with Field Data." *In preparation*.

[4] Land Information New Zealand (LINZ). (2018). "Digital Elevation Models and GIS Layers." Retrieved from https://data.linz.govt.nz/.

[5] Zhu, J., Daley, D., Baise, L.G., Baise, L.G, Thompson, E.M., Wald, D.J., and Knudsen, K.L. (2015). "A geospatial liquefaction model for rapid response and loss estimation." *Earthquake Spectra* 31(3): 1813-1837.

[6] Zhu, J., Baise, L.G., Thompson, E.M. (2017). "An updated geospatial liquefaction model for global application." *Bulletin of the Seismological Society of America* 107(3): doi: 10.1785/0120160198